On continuum approximations of discrete-state Markov processes of large system size

Davin Lunz

On continuum approximations of discrete-state Markov processes of large system size

Davin Lunz*

Abstract

Discrete-state continuous-time Markov processes are an important class of models employed broadly across the sciences. When the system size becomes large, standard approaches can become intractable to exact solution and numerical simulation. Approximations posed on a continuous state space are often more tractable and are presumed to converge in the limit as the system size tends to infinity. For example, an expansion of the master equation truncated at second order yields the Fokker–Planck equation, a widely used continuum approximation equipped with an underlying process of continuous state. Surprisingly, in [5] it is shown that the Fokker–Planck approximation may exhibit exponentially large errors, even in the infinite system-size limit. Crucially, the source of this inaccuracy has not been addressed. In this paper, we focus on the family of continuous-state approximations obtained by arbitrary-order truncations. We uncover how the exponentially large error stems from the truncation by quantifying the rapid error decay with increasing truncation order. Furthermore, we explain why this discrepancy only comes to light in a subset of problems. The approximations produced by finite truncation beyond second order lack underlying stochastic processes. Nevertheless, they retain valuable information that explains the previously observed discrepancy by bridging the gap between the continuous and discrete processes. The insight conferred by this broader notion of "continuum approximation", where we do not require an underlying stochastic process, prompts us to revisit previously expressed doubts regarding continuum approximations. In establishing the utility of higher-order truncations, this approach also contributes to the extensive discussion in the literature regarding the second-order truncation: while recognising the appealing features of an associated stochastic process, in certain cases it may be advantageous to dispense of the process in exchange for the increased approximation accuracy guaranteed by higher-order truncations.

1 Introduction

Discrete-state stochastic processes in general, and continuous-time Markov chains (also called Markov jump processes) in particular, are ubiquitous in physics, chemistry, biology, and throughout the applied sciences [8, 28]. The dynamics of these processes are described by the so-called master equation [9], a first-order system of ordinary differential equations (ODEs) governing the probability of the system to be in any given state at any given time.

In many realistic modeling scenarios, the complexity of a complete system description is beyond analytical tractability and computationally feasibility, driving the study of approximation techniques. There are several regimes in which the dynamics may be readily reduced, such as timescale separation [17, and references therein]. The regime we study in this work is when the system size (or population size) is very large. For example, in chemical reactions [24] the number of reactant molecules may be close to Avogadro's constant $\mathcal{O}(10^{23})$, while the number of proteins in a cell [23] can be on the order of $\mathcal{O}(10^6)$, and animal populations can be similarly large.

It is well established that "the relative fluctuations in the time-evolving species populations scale as the inverse square root of the reactant populations" [12]. This rule is often invoked to justify the use of a deterministic model to describe a stochastic process. It might thus be tempting to think that, in the limit of large system size, the influence of stochasticity ought to be neglected entirely. There are two important

*École Polytechnique, CMAP 91128 Palaiseau, France; Institut Pasteur, 75015 Paris, France; INRIA Saclay – Île de France, 91120 Palaiseau, France (davin.lunz@polytechnique.edu).

reasons why such a deterministic description is insufficient for our purposes. Firstly, in § 2 we study mean first-passage time problems, where we study the time it takes for the process to depart some distance from a metastable equilibrium state. This excursion from the metastable state eventually occurs (with probability one), even in the presence of arbitrarily weak stochasticity, as small fluctuations accumulate over sufficiently long transients. On the other hand, the deterministic dynamics do not admit any excursions from the metastable state. Mathematically, the large system-size limit and the large time limit do not commute, hence the deterministic systems are insufficient models in the study of these stochasticity-driven problems. Secondly, in practice, we are interested in being able to accurately describe systems of only moderately large size, where the noise might be asymptotically small but is certainly not to be neglected entirely, as will be evident in § 3.

The challenges associated with increasing system size have motivated various approximation methods (see [6, 10, 11, 27] and references therein). One common approach to exploit the large system size is to observe that, as the system size tends to infinity, each discrete jump makes a vanishingly small change in state, hinting that a continuous state may provide a suitable approximation setting. One may derive such a continuum approximation by expanding the master equation with respect to state, and truncating formally negligible terms to obtain a partial differential equation (PDE) describing the evolution of the probability density through space and time; a continuous-state analog of the master equation [12]. The lowest-order contribution describes the averaged drift, and the next order contribution describes the noise due to the stochasticity. When higher-order terms are neglected, the equation takes the form of a Fokker–Planck equation, which governs the evolution of the probability density of an underlying continuous-state process [26, Ch. 4].

For the sake of concreteness, we turn our attention to a specific subclass of Markov processes, known as birth–death processes. However, we emphasise that much of what we say may generalise to a broader class of models in higher dimension, as we will discuss. Birth–death processes are a prevalent subclass of Markov processes [5], describing the continuous-time dynamics of a system of discrete states enumerated by 0, 1, 2, …, Y (where Y can be ∞). Transitions only occur between neighbouring states in the enumeration, so $0 < X < Y$ can only transition to $X - 1$ or $X + 1$. The transition rate from state X to $X + 1$ is called the birth rate, and denoted by $\lambda(X)$, while the transition from X to $X - 1$ is the death rate, and denoted by $\mu(X)$.

The Fokker–Planck approximation has long been considered an accurate model for Markov processes in the large system-size limit [13]. Perhaps surprisingly, it was shown in [5] that the Fokker–Planck approximation of a birth–death process may deviate from the exact discrete solution substantially, producing exponentially large (absolute) errors, even as the system size tends to infinity. This observation was accompanied by "a heuristic argument" suggesting that the Fokker–Planck approximation is accurate only in the special circumstance that the birth and death rates remain sufficiently close for all states between the equilibrium state and the extinction state. It was thus concluded in [5] that the continuum approach is "delicate" and only applicable under very strict conditions, offering "a warning about the subtlety of the relationship between discrete and continuum approaches". It is worth noting that the mean first-passage time deviation of the Fokker–Planck approximation was reported previously in the literature [2, 20].

There are several steps in the derivation of the Fokker–Planck approximation from which error may originate. The transition from a discrete to continuum description of the population, alongside assuming some regularity of the continuum solution, are not trivial presumptions [12]. Non-analytic terms will not be captured by expanding the master equation [16]. The truncation neglects terms, although these are formally negligible in the large system-size limit. Previously, the precise source of the exponentially large approximation error was not identified, and therefore not remedied. In this paper, we argue that the continuum approximation (in a broad sense, as we will describe), even in enforcing additional structure than present in the discrete description, remains exponentially accurate. We present novel calculations showing that the errors stem from the finite truncation. The truncated terms, even though formally negligible, represent aspects of the discrete process whose effect over long time periods are, cumulatively, no longer vanishingly small. The calculations reveal that higher-order terms contribute successively less towards the first-passage time, and therefore a truncation remains reasonable, but might be required at order greater than two. We also analyse the stationary distribution of the process when extinction is excluded, and show that it possesses an analogous structure, however, due to normalisation of the distribution the influence of the higher-order terms is even less pronounced.

It is important to discuss the mathematical nature of the approximations under consideration. Truncating the expansion at any finite order beyond the second results in a PDE whose solution is not guaranteed to remain non-negative [25]. Therefore, the quantity governed by the equation may not be interpreted as a probability density, and thus such approximations are not associated with an underlying stochastic process. That the Fokker–Planck equation is equipped with an underlying Itō process confers enormous advantage by way of both complementary analytical approaches as well as numerical techniques based on stochastic sampling. More conceptually, the continuum model corresponding to the second-order truncation is a model in its own right, whereas higher order truncations cease to be independent models and serve only as density approximations. This presumably explains much of the popularity the second-order truncation enjoys.

Nevertheless, the value of higher-order approximations (even in the absence of sample trajectories) should not be overlooked. Even if their solution might become slightly negative, it remains a valid approximation, and may provide better overall predictive power (see [14] and [26, Ch. 4]). Positivity is an appealing property from a modeling perspective, however, in its absence, diminishing error at the cost of a favourable analytic property may be a profitable trade-off. Indeed, "all models are wrong the scientist must be alert to what is importantly wrong" [3].

In light of this, it is crucial to elucidate the notion of "continuum approximation" used in this work. We may be tempted to think that the discussion was motivated by seeking a continuous *process* approximating the original discrete process (and perhaps to which it converges, in some sense, as the system size tends to infinity). However, we emphasise that the key ingredient is a state continuum. Accordingly, we conceive of a "continuum approximation" to refer to any continuous-state approximation of the probability density of the discrete process, irrespective of an associated continuum-state process. The Fokker–Planck equation happens to correspond to a continuous-state process with all the accompanying advantages described. However, as we will describe, demanding this property imposes limitations on accuracy. Our broader conception invites the study of approximations that achieve greater accuracy (at the cost of abandoning an underlying process), and in so doing, sheds light on the connection between the continuous and discrete processes.

The rest of the paper is organised as follows. In § 2 we consider the problem of mean first-passage times. In § 2.1, we outline the previous results obtained in [5] for a particular birth–death process. In § 2.2 we describe the continuum approximation for arbitrary truncation order N, and demonstrate how a WKB approximation of these truncated systems exhibits vanishingly small error. In § 2.3 we compare the discrete and continuum descriptions as a function of the truncation order. In § 3 we study how the truncation considerations pertain to probability distributions. In § 3.1 we solve the discrete problem, in § 3.2 we analyse the Fokker–Planck and untruncated approximations, and quantify in § 3.4 how these converge to the discrete solution, highlighting how the higher-order terms are less consequential in this context. Finally, in § 4 we summarise our findings.

2 Mean first-passage time problems

2.1 Discrete birth–death processes

In this section we specify the birth–death processes of interest, and quote results from the literature, which will serve as a benchmark to which we compare the continuum approximations.

The birth–death process is described by the master equation, governing the probability P for the system to be in state X at time t, which is given by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}P(X,t) = \lambda(X-1)P(X-1,t) - \lambda(X)P(X,t) \\ + \mu(X+1)P(X+1,t) - \mu(X)P(X,t).\end{aligned} \tag{1}$$

Boundary conditions are also imposed. The state $X = 0$ typically represents population extinction at which the death rate is zero: $\mu(0) = 0$. Similarly, when the maximum state Y is finite, for example if the system has a finite capacity, the birth rate in state $X = Y$ must be zero: $\lambda(Y) = 0$ if $Y < \infty$. These boundary conditions guarantee the modeling description that only states $\{X\}_{X=0}^{Y}$ are populated (assuming initial conditions on this domain). It is typical to further assume that the birth rate at extinction is zero: $\lambda(0) = 0$, which makes $X = 0$ an absorbing state.

We are interested in the mean first-passage time to arrive at the extinction state $X = 0$ from any other state, which we denote by $T(X)$. The statistics of the time it takes to reach extinction are of great practical interest in several fields [5]. We highlight that, even without an absorbing state, we may study the first-passage time statistics to the $X = 0$ state, or any other state for that matter, and our results also pertain to these cases. An equation governing T may be deduced [28, Ch. 12], namely

$$-1 = \lambda(X)T(X+1) - [\lambda(X) + \mu(X)]T(X) + \mu(X)T(X-1), \tag{2}$$

for all processes where the state $X = 0$ is reached with probability 1 from all states. The second-order difference equation (2) is to be closed with two boundary conditions. For the sake of demonstration, we follow [5], and quote the results obtained there without reproducing the calculations. We consider a finite state space $0 \leq X \leq Y$, imposing an absorbing boundary conditions at $X = 0$ and a reflecting boundary condition at $X = Y$, which take the forms

$$T(0) = 0, \qquad T(Y) - T(Y-1) = 1/\mu(Y). \tag{3}$$

The system (2) and (3) admits the solution

$$T(X) = \sum_{i=1}^{X} \left[\frac{1}{\mu(i)} + \left(\prod_{j=1}^{i-1} \frac{\mu(j)}{\lambda(j)} \right) \sum_{k=i+1}^{Y} \frac{1}{\mu(k)} \prod_{\ell=1}^{k-1} \frac{\lambda(j)}{\mu(j)} \right]. \tag{4}$$

Our aim is to compare the solution (4) to the continuum approximation, for large system size, which we denote Ω. While the largest state in the model Y, is not, in principle, the same as the characteristic system size Ω, for the purposes of this paper, we do not distinguish between the two.

It is convenient to introduce scaled rates

$$\lambda(X) = \Omega\bar{\lambda}(X/\Omega), \qquad \mu(X) = \Omega\bar{\mu}(X/\Omega). \tag{5}$$

Following [5] we write $x = X/\Omega$ and consider the concrete case where

$$\bar{\lambda}(x) = \Lambda x(1-x), \qquad \bar{\mu}(x) = x, \tag{6}$$

corresponding to the discrete rates

$$\lambda(X) = \Lambda X(1 - X/\Omega), \qquad \mu(X) = X. \tag{7}$$

When $\Lambda > 1$ there is a metastable state $x = x_e := 1 - 1/\Lambda$ where $\bar{\lambda}(x_e) = \bar{\mu}(x_e)$. When $\Lambda < 1$ the origin is an attractor but there are no nonzero equilibria. The critical value $\Lambda = 1$ is therefore a threshold separating these two qualitatively distinct cases.

We focus exclusively on the superthreshold case where $\Lambda > 1$. Using the scalings (5) and the specific forms (6) in the exact solution (4), it is shown in [5] that, in the limit as $\Omega \to \infty$, the solution has a leading order asymptotic form $T(X) \sim T_e$ for $x = X/\Omega = \mathcal{O}(1)$ where

$$T_e = C\mathrm{e}^{\Omega\Phi(x_e)}, \tag{8}$$

for

$$C = \frac{1}{\bar{\lambda}(0) - \bar{\mu}(0)} \sqrt{\frac{2\pi\bar{\lambda}(0)\bar{\mu}(0)}{\Omega[\bar{\lambda}(x_e)\bar{\mu}'(x_e) - \bar{\lambda}'(x_e)\bar{\mu}(x_e)]}} + \mathcal{O}\left(\frac{1}{\Omega}\right), \tag{9}$$

$$\Phi(x) = \int_0^x \log\frac{\bar{\lambda}(\xi)}{\bar{\mu}(\xi)}\,\mathrm{d}\xi. \tag{10}$$

The exponent $\Omega\Phi(x_e)$ in (8) is positive and asymptotically large. Therefore, it is the central quantity of interest, and any deviation from this value in the continuum approach constitutes an exponentially large error. As such, our primary focus will be on $\log(T)/\Omega$, and, for the purposes of comparison with the continuum approximation, we note from (8) and (9) that

$$\log(T)/\Omega \sim \int_0^{x_e} \log\frac{\bar{\lambda}(\xi)}{\bar{\mu}(\xi)}\,\mathrm{d}\xi. \tag{11}$$

2.2 A family of continuum approximations

In an attempt to reduce system complexity, the continuum approach seeks to approximate the master equation (1) when the characteristic system size is large $\Omega \gg 1$. We adopt the normalised state $x = X/\Omega$ and the scalings (5), and introduce the scaled probability density p

$$p(x,t) = \Omega P(X,t). \tag{12}$$

Substituting p from (12) into the master equation (1), we consider $p(\cdot,t)$ to be defined on the continuum, which justifies a Taylor expansion of the master equation. Truncating the expansion at (possibly infinite) order N, we obtain

$$\frac{\partial}{\partial t}p_N(x,t) = \sum_{k=1}^{N} \frac{(-1)^k}{\Omega^{k-1}k!}\frac{\partial^k}{\partial x^k}\left[\bar{\lambda}(x)p_N(x,t)\right] + \sum_{k=1}^{N}\frac{1}{\Omega^{k-1}k!}\frac{\partial^k}{\partial x^k}\left[\bar{\mu}(x)p_N(x,t)\right]. \tag{13}$$

The Fokker–Planck equation is recovered by setting $N = 2$. We comment that there are two expansions employed in the literature to derive Fokker–Planck approximations: the Kramers–Moyal expansion and van Kampen's system size expansion. Our Taylor expansion is equivalent to the Kramers–Moyal expansion, and we refer the reader to [9, 14, 12] for further discussions in this direction.

We proceed to pose the problem of the mean first-passage time to $x = 0$ for densities governed by (13) starting from any state x, which we denote $\tau_N(x)$. As noted in § 1, strictly speaking, p_N is not a density for finite $N > 2$ and thus a first-passage time is not well-defined. Nonetheless, we proceed analogously for arbitrary N, bearing in mind that these do not describe an underlying process for finite $N > 2$ but merely approximate one. By taking the adjoint of the spatial operator on the right-hand side of (13), it may be shown [28, Ch. 12] that τ_N satisfies the ODE

$$\begin{aligned} -1 &= \bar{\lambda}(x)\sum_{k=1}^{N}\frac{1}{\Omega^{k-1}k!}\frac{\mathrm{d}^k}{\mathrm{d}x^k}\tau_N(x) + \bar{\mu}(x)\sum_{k=1}^{N}\frac{(-1)^k}{\Omega^{k-1}k!}\frac{\mathrm{d}^k}{\mathrm{d}x^k}\tau_N(x) \\ &= \sum_{k=1}^{N}\frac{a_k(x)}{\Omega^{k-1}k!}\frac{\mathrm{d}^k}{\mathrm{d}x^k}\tau_N(x), \end{aligned} \tag{14}$$

where

$$a_k(x) = \begin{cases} \bar{\lambda}(x) - \bar{\mu}(x), & k \text{ odd}, \\ \bar{\lambda}(x) + \bar{\mu}(x), & k \text{ even}. \end{cases} \tag{15}$$

Many previous studies avoid the continuum approach, citing the inaccuracy of the Fokker–Planck approximation, and apply a WKB approximation [15, Ch. 7.5] directly to the master equation [1, 4, 20, 22, 29]. We remain within the continuum framework, and solve (14) asymptotically, for $\mathrm{d}\tau_N/\mathrm{d}x$, in the large-system limit as $\Omega \to \infty$ by means of a WKB expansion, whereby

$$\frac{\mathrm{d}}{\mathrm{d}x}\tau_N(x) \sim \left(c_N(x) + \cdots\right)\mathrm{e}^{\Omega V_N(x)}. \tag{16}$$

Integrating from $x = 0$ and imposing the boundary condition $\tau_N(0) = 0$, we can write

$$\tau_N(x) = \int_0^x \frac{\mathrm{d}}{\mathrm{d}\xi}\tau_N(\xi)\,\mathrm{d}\xi. \tag{17}$$

It is important to observe from (16) that the dominant contribution from $\mathrm{d}^k\tau_N/\mathrm{d}x^k$ is given by

$$\frac{\mathrm{d}^k}{\mathrm{d}x^k}\tau_N(x) \sim c_N(x)\left[\Omega V_N'(x)\right]^{k-1}\mathrm{e}^{\Omega V_N(x)}, \tag{18}$$

where the prime denotes differentiation. Since the magnitude of the kth derivative is $\mathcal{O}\left(\Omega^{k-1}\right)$, we deduce that, despite each successive term being $\mathcal{O}(\Omega)$ smaller in the approximation (13), their contribution to the

exponentially large mean first-passage times are all the same order of magnitude, as $\Omega \to \infty$. Therefore, at leading order, the ODE (14) reduces to the form

$$0 = \sum_{k=1}^{N} \frac{a_k(x)}{k!} V_N'(x)^{k-1}. \tag{19}$$

From (19) and the specific forms (6), it follows that the even-indexed coefficients $a_{2k}(x) > 0$ for all x, while the odd-indexed coefficients $a_{2k-1}(x) > 0$ for all $x \in [0, x_e)$. Therefore, $V_N'(x) < 0$ for all $x \in [0, x_e)$ in order for the right-hand side of (19) to be zero. Since V_N' is negative for all $x \in [0, x_e)$, it holds that, on this same domain, $V_N(x_e) < V_N(x)$. Since x_e is a metastable equilibrium point, we expect the mean first-passage time to increase exponentially throughout this domain, except perhaps in a vicinity of the equilibrium point. Therefore, we expect $V_N(x) > 0$ except perhaps near $x \approx x_e$. This motivates us to choose the constant of integration such that $V_N(x_e) = 0$, namely

$$V_N(x) = \int_{x_e}^{x} V_N'(\xi)\, \mathrm{d}\xi. \tag{20}$$

Note that this choice is without loss of generality, since any other constant may be absorbed into the coefficient c_N defined in (16). Finally, from the previous observation that $V_N' < 0$ on $x \in [0, x_e)$, we may apply Laplace's method [7, Ch. 2] to the integral (17), to deduce that

$$\log(\tau_N)/\Omega \sim V_N(0) = -\int_{0}^{x_e} V_N'(\xi)\, \mathrm{d}\xi, \tag{21}$$

where τ_N denotes the mean first-passage time governed by the N-truncated approximation. The calculation is detailed in appendix A.

With (11) and (21) in mind, our aim is to show that the sequence V_N' converges to $-\Phi'$ given in (10), and therefore, the continuum approximations (beyond Fokker–Planck) converge to an exponentially accurate estimate of the discrete process.

We see from (19) that $V_N'(x)$ is a real root of a polynomial of degree $(N-1)$. For the Fokker–Planck case, where $N = 2$, we recover

$$V_2'(x) = -2\,\frac{\bar{\lambda}(x) - \bar{\mu}(x)}{\bar{\lambda}(x) + \bar{\mu}(x)}. \tag{22}$$

Considering first the case of $N = \infty$, we find from (19) that

$$\begin{aligned} 0 &= (\bar{\lambda}(x) - \bar{\mu}(x)) \sum_{k=1}^{\infty} \frac{1}{(2k-1)!} V_\infty'(x)^{2k-1} + (\bar{\lambda}(x) + \bar{\mu}(x)) \sum_{k=1}^{\infty} \frac{1}{(2k)!} V_\infty'(x)^{2k} \\ &= \left[\bar{\lambda}(x) - \bar{\mu}(x)\right] \sinh V_\infty'(x) + \left[\bar{\lambda}(x) + \bar{\mu}(x)\right] (-1 + \cosh V_\infty'(x)), \end{aligned} \tag{23}$$

where we have identified the hyperbolic trigonometric power series. Neglecting the trivial solution, equation (23) is solved by

$$V_\infty'(x) = \log \frac{\bar{\mu}(x)}{\bar{\lambda}(x)}. \tag{24}$$

2.3 Comparing discrete and continuum mean first-passage times

In the Fokker–Planck case of $N = 2$, we substitute (22) into (21) to find that

$$\log(\tau_2)/\Omega \sim 2 \int_{0}^{x_e} \frac{\bar{\lambda}(\xi) - \bar{\mu}(\xi)}{\bar{\lambda}(\xi) + \bar{\mu}(\xi)}\, \mathrm{d}\xi, \tag{25}$$

in agreement with [5].

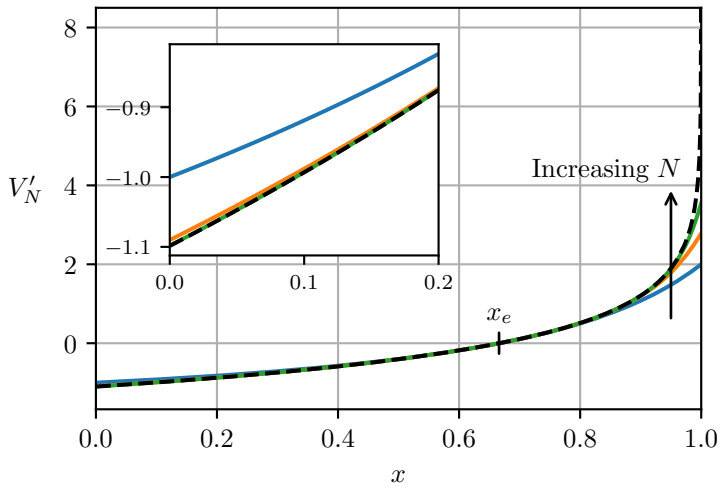


Figure 1: Plots of $V'_N(x)$ for $N \in \{2, 4, 6, \infty\}$. The convergence to the limit of $N = \infty$, plotted as the black dashed curve, is very rapid, near the diverging endpoint $x \approx 1$. The inset shows a zoomed in section of the plot around the endpoint $x \approx 0$.

In the case of $N = \infty$, we find from (21) and (24) that

$$\log(\tau_\infty)/\Omega \sim \int_0^{x_e} \log \frac{\bar{\lambda}(\xi)}{\bar{\mu}(\xi)} \, \mathrm{d}\xi, \tag{26}$$

which coincides with the asymptotics of the exact solution (11).

We now address intermediate values of $2 < N < \infty$. Since the series expressions in (23) have an infinite radius of convergence, we expect these to converge to the value $\log[\bar{\mu}(x)/\bar{\lambda}(x)]$ for all x for which the logarithm is finite. Moreover, on any domain in which the logarithm is uniformly bounded, we expect the convergence to be uniform with respect to x. In particular, in the case given by (6), we expect uniform convergence over $x \in (0, 1 - \delta)$ for any $\delta > 0$.

Naturally, the convergence requires sufficiently large values of N, however, for small values of N the approximates V'_N may not be well-defined over the entire domain. For example, when $N = 3$, we may solve the quadratic (19) to find that

$$V'_3(x) = \frac{-3\left[\bar{\lambda}(x) + \bar{\mu}(x)\right] \pm \sqrt{3}\sqrt{12\bar{\lambda}(x)\bar{\mu}(x) - 5[\bar{\lambda}(x) - \bar{\mu}(x)]^2}}{2[\bar{\lambda}(x) - \bar{\mu}(x)]}. \tag{27}$$

Only one of these roots closely approximates V'_∞ in the domain, however, the more serious problem is that V'_3 is not real for all $x \in [0, 1]$. We can see that, for the forms (6), for $x \approx 1$ we have $\bar{\lambda}(x) \approx 0$, so the first term inside the larger square-root will be arbitrarily small, while the second term is approximately $-5\bar{\mu}(x)^2 \approx -5 < 0$. Our numerical results show that this issue does not arise for even values of N, while for odd values of N it is relegated to increasingly smaller regions as N increases, as expected.

To demonstrate this convergence numerically, we use the particular forms in (6), and choose $\Lambda = 3$, since this is the value chosen in [5] to demonstrate the "failure" of Fokker–Planck. We determine $V'_N(x)$ by numerical root-finding of equation (19).

First, in fig. 1, we show how the iterates $V'_N(x)$, for even N, rapidly converge to the black dashed limiting curve V'_∞. We see that, as highlighted in [5], the discrepancy of the Fokker–Planck approximation $N = 2$ is noticeable. However, already by $N = 4$ this discrepancy is barely visible over the domain of primary interest $x \in [0, x_e]$. The discrepancy is still visible near the endpoint $x \approx 1$ as V'_∞ diverges, however, this error is of secondary importance since the value of the mean first-passage time is set by V'_N on $x \in [0, x_e]$, as seen in (21).

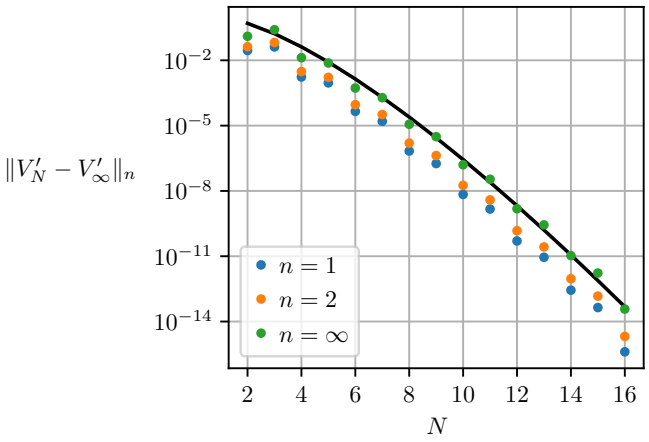


Figure 2: The $L^n([0, 0.9])$ norm of the approximation error of V'_N for increasing N. The black curve shows the predicted convergence behaviour, $1/N!$.

Second, we look to quantify the convergence. We expect the truncated series in (23) to have an error of order $\mathcal{O}(1/N!)$ as $N \to \infty$ on any domain where V'_∞ is bounded, say $x \in [0, 1-\delta]$ for any $\delta > 0$. In fig. 2 we plot the error with $\delta = 0.1$, to find that the convergence matches the predicted rate.

Recall that the odd iterates may not exist everywhere. While all odd iterates do exist on the domain $x \in [0, 1-\delta]$ for $\delta = 0.1$ and $\Lambda = 3$, we see that, in particular for small values of N, the error drops much more significantly for even N than for odd N. In fact, the error increases from $N = 2$ to $N = 3$ for this choice of parameters.

In addition to the large system-size limit of $\Omega \to \infty$, we have introduced the new limit of increasing truncation order $N \to \infty$. The asymptotic nature of the limit $\Omega \to \infty$ means that, for large but finite Ω, there might be an optimal truncation point [26, Ch. 4]; some finite N_0 beyond which the error does not decrease for increasing truncation order $N > N_0$.

To recap, we analyse the logarithm of the mean first-passage time for continuum approximations truncated at order N. We find that these approximations converge rapidly to the exact solution with increasing truncation order. While [5] concludes from the discrepancy of the Fokker–Planck approximation of $N = 2$ that the continuum approach is delicate, this analysis offers a different perspective. Even though the low-order truncation of $N = 2$ exhibits some discrepancy in the log mean first-passage time, it certainly captures the qualitative behaviour. Moreover, this discrepancy is significantly reduced for higher truncation order, with the error decreasing superexponentially as $\mathcal{O}(1/N!)$ for large N. This analysis illuminates the nature of the discrepancy between the continuous-state process described by the Fokker–Planck equation, and the discrete-state process governed by the master equation: the truncated terms bridge the gap between the underlying continuous and discrete processes, and the truncation order acts as a (discrete) interpolation between the two.

We finish by highlighting that these calculations were performed for fairly generic birth and death rates, $\bar{\lambda}$ and $\bar{\mu}$, respectively. The particular forms (6) were used merely for the purposes of illustration. Moreover, we did not specifically need that the first-passage be sought to an absorbing state. The single crucial ingredient is that the solution of the first-passage problem be of WKB form (16) (with $\mathcal{O}(1)$ coefficient and exponent), which leads to higher-order terms in the expansion making $\mathcal{O}(1)$ contributions to the first-passage times. This occurs for processes with a metastable state (or several), and a first-passage to any other state separated from the equilibrium by $\mathcal{O}(\Omega)$ discrete states. Physically, the exponentially long first-passage times are born out of the metastability in combination with small fluctuations: to reach the threshold state requires the rare accumulation of small fluctuations. While these fluctuations are vanishingly small, over a sufficiently large time interval they can (and will) lead to the excursion from the metastable state, and are thus not at all negligible. We deduce that the continuum approach provides a systematic and robust family of approximations for one-dimensional birth–death processes with metastable equilibria. While more

conjectural, it seems that this analysis could extend to more exotic Markov processes beyond one dimension and the birth–death structure, such as those constructed in [19].

3 Extinctionless stationary distributions

The purpose of this paper is to affirm the usefulness and accuracy of the continuum approach of approximating discrete Markov processes. Having shown that, for mean first-passage time problems, the exponential error of the Fokker–Planck approximation is recoverable by a higher truncation order, we now analyse the accuracy of the continuum approximations for the problem of distributions governed by the master equation (1). While a similar exponential structure is exposed in this problem, we find that its influence on the distributions is exponentially smaller due to the distributions being normalised. This solidifies the earlier explanation — that the discrepancy observed in [5] is due to the exponentially large passage times, when the accumulation of smaller effects may not be negligible. We further demonstrate that the continuum approximations are remarkably accurate, even for very moderate system sizes.

The simplest comparison is between stationary distributions. The unique stationary distribution of the birth–death process using the rates in (6) is extinction. A more interesting distribution is the quasistationary distribution (QSD): that to which the process settles on intermediate timescales, longer than the birth and death rates characterising the transitions, but shorter than the exponentially long timescales at which extinction is relevant. The QSD is formally defined via the conditional distribution of the Markov process at time t, conditioned on extinction not having occurred, by taking the limit as $t \to \infty$ [21]. This distribution is typically studied via the spectrum of the generator matrix associated with the master equation, or the spectrum of the differential operator associated with the Fokker–Planck equation. The single zero eigenvalue is associated with the eigenvector corresponding to the unique steady state. The remaining eigenvalues are negative, and the eigenvector associated with that of smallest magnitude corresponds to the QSD. Appealing to spectral theory is powerful in that it provides not just these limiting distributions, but also transient information regarding the rates of convergence. However, closed-form expressions for a general eigendecomposition are unattainable, which prohibits the comparison we seek to perform. Instead, we choose to study the stationary distribution of a modified process where extinction is excluded (rather than conditioning on not having occurred), which we call the extinctionless stationary distribution (ESD).

3.1 The discrete ESD

To define the ESD of the discrete process, we begin by considering a modified process where extinction is excluded. This may be achieved by setting the birth rate at the state $X = 0$ to be positive, that is, $\lambda(0) = \epsilon > 0$, and seeking the unique stationary distribution of this modified process.

The stationary distribution may be obtained by setting the left-hand side of the master equation (1) to zero, whereby

$$\begin{aligned}
0 &= -\lambda(0)P(0) + \mu(1)P(1),\\
0 &= \lambda(0)P(0) - (\lambda(1) + \mu(1))P(1) + \mu(2)P(2),\\
&\vdots\\
0 &= \lambda(Y-2)P(Y-2) - (\lambda(Y-1) + \mu(Y-1))P(Y-1) + \mu(Y)P(Y),\\
0 &= \lambda(Y-1)P(Y-1) - \mu(Y)P(Y).
\end{aligned} \tag{28}$$

Successive substitution in (28) shows that $\lambda(X)P(X) = \mu(X+1)P(X+1)$ for all $0 \leq X \leq Y-1$, from which we deduce that, for all $X > 0$,

$$P(X) = \lambda(0)P(0)\frac{1}{\mu(X)}\prod_{i=1}^{X-1}\frac{\lambda(i)}{\mu(i)}, \tag{29}$$

which determines $P(X)$ for all states $X > 0$. To determine $P(0)$, note that the system of recurrence relations (28) is homogeneous in P, and therefore any solution may be multiplied by a constant to yield

another solution. For a valid distribution, we impose $\sum_{X=0}^{Y} P(X) = 1$ which resolves this indeterminacy, and proves that the stationary distribution is unique.

Observing that $P(0) = \mathcal{O}(1)$ while $P(X) = \mathcal{O}(\epsilon)$ for all $X > 0$, we deduce that, in the limit as $\epsilon \to 0$, the stationary distribution concentrates all probability mass in the state $X = 0$; in other words, extinction, as to be expected. Since we are not interested in the extinction state, we neglect the $X = 0$ state by defining the distribution shared among the remaining states $X > 0$, which we denote $P_{X>0}(X)$, and see from (29) that $P_{X>0}(X)$ is given by

$$P_{X>0}(X) = \frac{C}{\mu(X)} \prod_{i=1}^{X-1} \frac{\lambda(i)}{\mu(i)}, \qquad \frac{1}{C} = \sum_{X=1}^{Y} \frac{1}{\mu(X)} \prod_{i=1}^{X-1} \frac{\lambda(i)}{\mu(i)}. \tag{30}$$

We refer to the distribution $P_{X>0}$ in (30) as the ESD. Another way of deriving (30) is to consider the leading-order distribution in the asymptotic limit as $\epsilon \to \infty$, which corresponds to replacing the absorbing boundary with a reflecting boundary.

For the purposes of later comparison, we now deduce an approximate form for the ESD. In [5], it is shown that, in the large system limit $\Omega \gg 1$,

$$\prod_{i=1}^{X-1} \frac{\lambda(i)}{\mu(i)} \sim \frac{\sqrt{\bar{\mu}(0)\bar{\mu}(x)}}{\sqrt{\bar{\lambda}(0)\bar{\lambda}(x)}} \, \mathrm{e}^{\Omega\Phi(x)}, \tag{31}$$

where Φ is defined in (10) and we recall that $x = X/\Omega$. Therefore,

$$P_{X>0}(X) \sim \frac{\bar{C}}{\sqrt{\bar{\lambda}(x)\bar{\mu}(x)}} \, \mathrm{e}^{\Omega\Phi(x)}, \tag{32}$$

having absorbed all constant terms into the normalising constant $\bar{C}$.

Since $\Omega \gg 1$, the term on the right-hand side of (32), and therefore the ESD, is exponentially dominated by regions where the $\Phi(x)$ takes its maximum value. We note that, with the specific forms (6), the integrand in $\Phi(x)$ takes the form $\log[\bar{\lambda}(x)/\bar{\mu}(x)] = \log[\Lambda(1-x)]$, and is therefore positive on $x \in [0, x_e)$ and negative on $x \in (x_e, 1]$. Therefore $\Phi(x)$ has a unique maximum value at $x = x_e$, and we approximate the ESD via a second-order Taylor expansion, from which we find that the ESD is described by the Gaussian

$$P_{X>0}(X) \approx c \, \mathrm{e}^{-\Omega\Lambda(x-x_e)^2/2}, \qquad \frac{1}{c} = \int_0^1 \mathrm{e}^{-\Omega\Lambda(\xi-x_e)^2/2} \, \mathrm{d}\xi. \tag{33}$$

Equipped with the exact discrete ESD and an approximate form, we proceed to derive the ESD of the continuum approximations.

3.2 The Fokker–Planck ESD

We begin the continuum approximation of the extinctionless stationary distribution by writing down the Fokker–Planck equation, obtained by setting $N = 2$ in (13), namely

$$\frac{\partial}{\partial t} p_2(x,t) = -\frac{\partial}{\partial x} \left[\left(\bar{\lambda}(x) - \bar{\mu}(x) \right) p_2(x,t) \right] + \frac{1}{2\Omega} \frac{\partial^2}{\partial x^2} \left[\left(\bar{\lambda}(x) + \bar{\mu}(x) \right) p_2(x,t) \right]. \tag{34}$$

The associated boundary conditions are given by imposing conservation [8, Ch. 5], whereby

$$\left(\bar{\lambda}(x) - \bar{\mu}(x) \right) p_2(x,t) - \frac{1}{2\Omega} \frac{\partial}{\partial x} \left[\left(\bar{\lambda}(x) + \bar{\mu}(x) \right) p_2(x,t) \right] = 0, \tag{35}$$

at the boundary points. With the specific rates (6), the problem (34) and (35) governing p_2 takes the form

$$\frac{\partial}{\partial t} p_2(x,t) = -\frac{\partial}{\partial x} \left[(\Lambda x(1-x) - x) \, p_2(x,t) \right] + \frac{1}{2\Omega} \frac{\partial^2}{\partial x^2} \left[(\Lambda x(1-x) + x) \, p_2(x,t) \right], \tag{36}$$

subject to the boundary conditions

$$(\Lambda x(1-x)-x)\,p_2(x,t) - \frac{1}{2\Omega}\frac{\partial}{\partial x}\left[(\Lambda x(1-x)+x)p_2(x,t)\right] = 0 \text{ at } x = 0, 1. \tag{37}$$

Note that the no-flux boundary condition at the left-hand boundary $x = 0$ corresponds to the reflecting limit discussed for the discrete case in § 3.1. However, in the case of degenerate diffusion at the boundary, this is formally equivalent to an absorbing boundary condition, as is the case in (37).

We now examine the Fokker–Planck problem (34) and (35), for which we present exact solutions in the steady case. We also discuss asymptotic approximations, which provide explicit expressions for the transient dynamics, and reveal the underlying structure of the solutions which aids a subsequent convergence analysis.

3.2.1 Exact steady solution

As in the discrete case, an exact solution of (34) and (35), which we denote by p_2, is available in the steady case. Neglecting the time derivative, and integrating (34) once shows that the flux is constant. Then the no-flux boundary conditions (35) are satisfied at both boundaries, and we are left with the first-order ODE

$$\left(\bar{\lambda}(x) - \bar{\mu}(x)\right) p_2(x) - \frac{1}{2\Omega}\frac{\mathrm{d}}{\mathrm{d}x}\left[\left(\bar{\lambda}(x) + \bar{\mu}(x)\right) p_2(x)\right] = 0. \tag{38}$$

Writing $q = \left(\bar{\lambda} + \bar{\mu}\right) p_2$, we see from (38) that

$$\frac{\mathrm{d}}{\mathrm{d}x}\log(q(x)) = \frac{q'(x)}{q(x)} = 2\Omega\,\frac{\bar{\lambda}(x) - \bar{\mu}(x)}{\bar{\lambda}(x) + \bar{\mu}(x)}, \tag{39}$$

from which it follows that

$$p_2 = \frac{\hat{c}_2\,\mathrm{e}^{\Omega W_2(x)}}{\bar{\lambda}(x) + \bar{\mu}(x)} = \frac{\hat{c}_2\,\mathrm{e}^{\Omega W_2(x)}}{\Lambda x(1-x) + x}, \tag{40}$$

for a normalising constant $\hat{c}_2$, and

$$W_2(x) = 2\int \frac{\bar{\lambda}(x) - \bar{\mu}(x)}{\bar{\lambda}(x) + \bar{\mu}(x)}\,\mathrm{d}x = 2\left(x + \frac{2}{\Lambda}\log[1 + \Lambda(1-x)]\right) + \text{const.} \tag{41}$$

Note that W_2 in (41) is none other than $-V_2$ found in the first-passage time problem (22).

The expression (40) with the specific rates (6) diverges as $x \to 0$, due to an accumulation of probability mass at the absorbing origin. Moreover, p_2 asymptotes like $p_2 \sim 1/x$ as $x \to 0$, which is not integrable on $x \in [0, 1]$. Therefore, $p_2(x)$ is not formally a distribution. The non-integrability stems from the degenerate diffusion: the unsteady problem (34) and (35) is conservative, with $\int_0^1 p_2(x,t)\,\mathrm{d}x = 1$ for all t, but singular at $x = 0$. Therefore, while a boundary layer exists at $x = 0$ for all time, the calculation of (40) reflects a pointwise (but non-uniform) convergence where the boundary layer has disappeared, leading to a divergent integral. As explored in § 2, we expect this accumulation to be relevant only on exponentially long timescales. Our interest in intermediate timescales suggests that the steady form (40) contains a wealth of information if we suitably neglect a neighbourhood of the origin. We pursue this by considering the normalised solution on the subdomain $[\epsilon, 1]$, for $\epsilon \ll 1$, which removes the divergent contribution in analogy with the exclusion of extinction in the discrete probability (30). For $\epsilon \ll 1$ not too small (as we will describe) this remains an accurate description of the ESD since the drift steers the process away from the origin, thereby ensuring that the divergent contribution is exponentially localised, as we now demonstrate. With this regularisation in mind, we continue to refer to p_2 as a distribution.

The factor $W_2(x)$ in the exponent in (40) obtains a maximum at the equilibrium point $x = x_e = 1 - 1/\Lambda$. If $\epsilon \gg \mathrm{e}^{-\Omega[W_2(x_e) - W_2(\epsilon)]}$, which is exponentially small since $W_2(x_e) - W_2(0) = \mathcal{O}(1)$ in the limit $\Omega \to \infty$, then $p_2(x_e) \gg p_2(\epsilon)$. This guarantees that the exact solution p_2 on the perturbed domain is dominated by the contributions away from the origin, whereby the divergent accumulation is suppressed and p_2 closely approximates the ESD. The normalisation constant $\hat{c}_2$ is then given by

$$\frac{1}{\hat{c}_2} = \int_\epsilon^1 \frac{\mathrm{e}^{\Omega W_2(\xi)}}{\Lambda\xi(1-\xi) + \xi}\,\mathrm{d}\xi. \tag{42}$$

The exact solution p_2 in (40) is explicit, up to the constant $\hat{c}_2$ defined in (42), which may be evaluated by numerical quadrature. Care must be taken to choose the constant of integration for W_2 such that W_2 does not become positive, whereby the calculation becomes susceptible to large errors and numerical overflow, as the integrand extends beyond the range available for standard floating-point calculation. An asymptotic approximation of $\hat{c}_2$ via Laplace's method (detailed in appendix A), produces a closed-form expression that alleviates these problems, namely

$$\begin{aligned}\hat{c}_2 &\sim [\bar{\lambda}(x_e) + \bar{\mu}(x_e)]\, \mathrm{e}^{-\Omega W_2(x_e)} \sqrt{\frac{-\Omega W_2''(x_e)}{2\pi}} \\ &= 2(1 - 1/\Lambda)\, \mathrm{e}^{-\Omega W_2(x_e)} \sqrt{\frac{\Omega\Lambda}{2\pi}}.\end{aligned} \tag{43}$$

Combining the exponents in (40) and (43) we see that $p_2 \propto \mathrm{e}^{\Omega[W_2(x) - W_2(x_e)]}$, which is no longer exponentially large. In fact, this expression shows that to integrate the exponent from $x = x_e$ is the natural choice of constant, just as discussed in the context of first-passage time problems in the paragraph containing (20).

It might be tempting to think that once we have an exact solution there is little more to do. However, the exact solution was only for the steady problem. Moreover, in higher dimensional problems, exact solutions may not be available. For these reasons, we proceed to explore asymptotic solutions. As it turns out, these provide additional quantitative insight even for the steady problem when we have an exact solution.

3.2.2 Asymptotic solution

In this section, our aim is to solve (36) and (37) asymptotically, in the limit as $\Omega \to \infty$. To this end, we expand p_2 asymptotically in inverse powers of Ω

$$p_2 \sim p_{2,0} + \frac{1}{\Omega} p_{2,1} + \cdots. \tag{44}$$

The leading-order form of the governing PDE (36) takes the form

$$\frac{\partial}{\partial t} p_{2,0}(x,t) + \frac{\partial}{\partial x} \left[(\Lambda x(1-x) - x)\, p_{2,0}(x,t) \right] = 0, \tag{45}$$

subject to

$$p_{2,0}(0,t) = p_{2,0}(1,t) = 0. \tag{46}$$

The hyperbolic PDE (45) and (46) may be solved by the method of characteristics. Characteristic curves satisfy

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \Lambda x(1-x) - x, \tag{47}$$

which admits the general solution

$$x = \left[\frac{1}{x_e} + \left(\frac{1}{x_0} - \frac{1}{x_e} \right) \mathrm{e}^{-(\Lambda-1)(t-t_0)} \right]^{-1}, \tag{48}$$

where the characteristic passes through $x = x_0$ at $t = t_0$. Along characteristics, the density satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t} p_{2,0}(t) = [2\Lambda x(t) - (\Lambda - 1)]\, p_{2,0}(t). \tag{49}$$

Substituting the characteristic form of $x(t)$ from (48) and integrating, we find that, along characteristics, the probability satisfies

$$p_{2,0}(t) = p_{2,0}(t_0) \mathrm{e}^{(\Lambda-1)(t-t_0)} \left[\frac{x_0}{x_e} + \left(1 - \frac{x_0}{x_e} \right) \mathrm{e}^{-(\Lambda-1)(t-t_0)} \right]^{2\Lambda}. \tag{50}$$

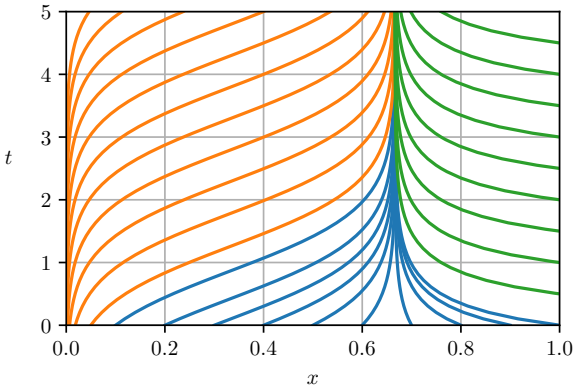


Figure 3: Characteristics (48) for $\Lambda = 3$, emanating from $(x_0, t_0) \in \cup_{i=1}^{3} X_i \times T_i$, where $X_1 = \{j/10\}_{j=1}^{10}$, $X_2 = \{\mathrm{e}^{-j}\}_{j=3}^{13}$, $X_3 = \{1\}$, $T_1 = T_2 = \{0\}$, and $T_3 = \{j/2\}_{j=1}^{9}$. The blue, orange, and green characteristics are associated with $X_i \times T_i$ for $i = 1, 2, 3$, respectively.

For any fixed characteristic with $x_0 > 0$, the term in (50) in square brackets is bounded, and we highlight the fact that the probability density grows exponentially.

We illustrate typical characteristic curves on the (x, t)-plane in fig. 3 for $\Lambda = 3$. We observe that all the characteristics converge towards the equilibrium point $x = x_e = 1 - 1/\Lambda = 2/3$. From (48) we deduce that the convergence is exponential, and from (49) we find that the total density is conserved by exponential growth of the probability density along characteristics. Thus, for any initial condition, the leading-order probability density tends to a delta function, with all the probability mass being concentrated at $x = x_e$.

The leading-order boundary conditions are $p_{2,0} = 0$ at both $x = 0$ and $x = 1$, thus no probability mass enters the domain from the boundaries. This is less important at the left-hand boundary $x = 0$ where the advection vanishes, as illustrated by the orange characteristics that fill the entire upper-left side of the domain, despite all emanating from the x-axis at $t = 0$. However, the green characteristic curves show that the zero boundary condition at $x = 1$ does enter the domain.

The convergence of the leading-order solution towards a delta function does not agree with the discrete ESD (33). The discrepancy is not a matter of proceeding to higher order, since the contribution of higher-order terms to the steady state is beyond all orders. Instead, an inner (boundary) layer of width $\mathcal{O}(1/\sqrt{\Omega})$ develops around $x = x_e$, in which the higher-order terms are not negligible. We may resolve the probability density in the vicinity of this inner layer by adopting the scalings

$$x = x_e + \xi/\sqrt{\Omega}, \qquad p(\xi) = p_{2,0}(x)/\sqrt{\Omega}. \tag{51}$$

Within this inner layer, the equation (36), along with the reinstated second-order terms, takes the form

$$\frac{\partial p}{\partial t} = (\Lambda - 1)\frac{\partial}{\partial \xi}(\xi p) + (1 - 1/\Lambda)\frac{\partial^2 p}{\partial \xi^2}, \tag{52}$$

up to terms of order $\mathcal{O}(1/\sqrt{\Omega})$, and we impose the matching conditions $p \to 0$ as $\xi \to \pm\infty$. Since we only obtained the stationary distribution in the discrete case, it suffices to solve the steady inner problem, which admits the solution

$$p = c\,\mathrm{e}^{-\Lambda\xi^2/2}, \tag{53}$$

where c is a normalising constant. Rewriting (53) in terms of the outer probability density by undoing the scalings (51), we find that

$$p_{2,0} = c_{2,0}\,\mathrm{e}^{-\Omega\Lambda(x-x_e)^2/2}, \qquad c_{2,0} \sim \sqrt{\frac{\Omega\Lambda}{2\pi}}, \tag{54}$$

where $c_{2,0} = c\sqrt{\Omega}$ remains a normalising constant, admitting an asymptotic approximation via Laplace's method (see appendix A). Since the outer problem is zero beyond the inner layer, the solution (54) is uniformly valid in x. The Fokker–Planck approximation (54) matches the Taylor-expanded form of the discrete solution (33).

The asymptotic analysis reveals the time-dependent convergence towards the equilibrium, as well as the structure around the equilibrium. The asymptotic structure of the outer and inner solutions shows that the deterministic dynamics are dominant beyond the region of width $\mathcal{O}(1/\sqrt{\Omega})$ around the equilibrium, wherein the stochastic dynamics are non-negligible. This scaling will be crucial in estimating the approximation error, not just of the asymptotic solution but also of the exact solution and higher-order solution.

3.3 Higher-order truncations

Both discrete approximations (32) and (33), the exact solution (40), and the asymptotic approximation (54), are all of the form $c(x)\,\mathrm{e}^{\Omega W(x)}$, for various c and W. This suggests that a WKB approximation might be helpful. Indeed, our earlier analysis in § 2 begs the question here: what effect is there on the distribution p_N when truncating the continuum approximation at higher order?

We take the WKB approximation for p_N (not its derivative, as was done in (16)) via

$$p_N \sim (c_N(x) + \cdots)\,\mathrm{e}^{\Omega W_N(x)}. \tag{55}$$

Substituting (55) into the steady version of equation (13), and proceeding just as in § 2.2, one finds that W_N satisfies

$$0 = \sum_{k=1}^{N} \frac{b_k(x)}{k!} W_N'(x)^{k-1}. \tag{56}$$

where $b_k = (-1)^k a_k$ defined in (15). This is analogous to (19), but since the equation governing p_N is the adjoint of that governing τ_N, a sign change in the odd coefficients results.

For Fokker–Planck, where $N = 2$, we recover $W_2 = -V_2$ as above, and proceeding to the next order we find that $c_2(x) = \hat{c}_2/\left[\bar{\lambda}(x) + \bar{\mu}(x)\right]$. That is, the WKB approximation recovers the exact solution.

In the case of $N = \infty$, and following the algebra from § 2, we find that

$$W_\infty' = \log\frac{\bar{\lambda}(x)}{\bar{\mu}(x)} = -V_\infty' = \Phi', \tag{57}$$

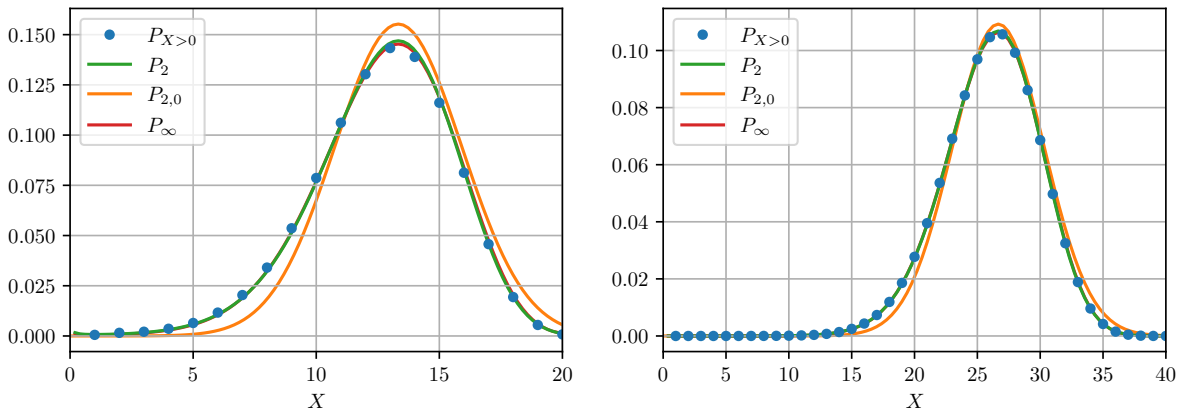
in precise agreement with the discrete approximation (31). Despite it being subdominant, for the sake of completeness we calculate $c_\infty(x)$ by proceeding to the next order (the calculations are performed in appendix B), giving the approximation

$$p_\infty \sim \frac{\hat{c}_\infty}{\sqrt{\bar{\lambda}(x)\bar{\mu}(x)}}\,\mathrm{e}^{\Omega\Phi(x)}, \qquad \hat{c}_\infty \sim (1 - 1/\Lambda)\,\mathrm{e}^{-\Omega\Phi(x_e)}\sqrt{\frac{\Omega\Lambda}{2\pi}}. \tag{58}$$

Just as for the Fokker–Planck case of $N = 2$, the function p_∞ is not integrable at the origin, where $p_\infty \sim 1/x$ as $x \to 0$. Despite there also being a singularity at $x = 1$, this is removable since $p_\infty \sim \sqrt{1-x}$ as $x \to 1$. Again we see from (58) that $x = x_e$ is the natural place to integrate from, however, we preserve the definition of Φ taken from [5].

3.4 Comparing discrete and continuum ESDs

Let us recap the expressions we determined for the ESD. Given the exact discrete solution $P_{X>0}$, we derived the closed-form approximation (32), and, Taylor expanding the exponent to second order, a further approximation (33). For the continuum family of approximations indexed by N, we obtained an exact Fokker–Planck ($N = 2$) solution (40), denoted p_2, an asymptotic Fokker–Planck solution (54), denoted $p_{2,0}$, and a WKB approximation of the $N = \infty$ case (58), denoted p_∞. In this section, we compare these expressions and draw conclusions by comparing and contrasting these results to the first-passage time results of § 2.



(a) $\Omega = 20$. (b) $\Omega = 40$.

Figure 4: A comparison of the ESDs. Exact discrete solution $P_{X>0}$ given by (30), the exact Fokker–Planck solution P_2 given in (40), the asymptotic Fokker–Planck solution $P_{2,0}$ given in (54), and the WKB approximation of the untruncated equation P_∞ given in (58). The comparison is done in the original scaling, hence the capital letters, by undoing (12).

The discrete approximation (32) matches exactly with p_∞ from (58). The Taylor expansion of the former (33) matches precisely with the asymptotic Fokker–Planck solution $p_{2,0}$ from (54). As with the first-passage time problem, despite being formally smaller, all higher orders contribute to the exponent. In contrast to the first-passage time case, here this discrepancy is particularly negligible. This is because the probabilities are normalised, therefore the probability density is exponentially small beyond a neighbourhood around the equilibrium point $x = x_e$, and local to the equilibrium, the exponents agree to second order. The outcome is that the errors associated with finite truncation are asymptotically small. We proceed to demonstrate this concretely by analysing the error asymptotics and demonstrating convergence, noting the quality of the approximations even for moderate system sizes Ω.

For the sake of measuring the accuracy of the approximations, we compare them to the exact discrete solution $P_{X>0}$ given in (30). To this end, we undo the scaling (12) and denote the probability densities in capital letters. In fig. 4, we illustrate the distribution for different moderately large values of Ω. The solutions P_2 and P_∞ agree remarkably well even for $\Omega = 20$, being almost indistinguishable. We see in fig. 4a good qualitative agreement for the asymptotic approximation $P_{2,0}$, and quantitatively improved agreement in fig. 4b. For smaller values of Ω the ESD deviates noticably from the Gaussian distribution. It is worth emphasising that the typical system size — a natural parameter choice to scale the system with — is more accurately given in these cases by $x_e\Omega = (2/3)\Omega$, which makes the agreement even more impressive.

From the matched asymptotic expansion in § 3.2.2 we deduce that $\mathcal{O}(\sqrt{\Omega})$ states in the vicinity of the equilibrium will contribute significantly to the probability density, while states beyond this region will be exponentially small. This allows us to analyse errors local to the equilibrium, which we do in three stages.

First, we compare the matched asymptotic Fokker–Planck solution $P_{2,0}$ to the exact Fokker–Planck solution P_2. The boundary-layer analysis performed to produce (54) is formally accurate up to $\mathcal{O}(1/\Omega)$. We thus predict that the L^n-norm of the discrepancy should asymptote like

$$\|P_{2,0} - P_2\|_n = \left(\sum_X [P_{2,0}(X) - P_2(X)]^n\right)^{1/n} = \mathcal{O}\left(\left[\sqrt{\Omega}\,\frac{1}{\Omega^n}\right]^{1/n}\right) = \mathcal{O}\left(\Omega^{-1+1/(2n)}\right). \tag{59}$$

Second, we compare the exact Fokker–Planck solution P_2 and the WKB approximation of the untruncated system P_∞, taking only contributions from $|x - x_e| = \mathcal{O}(1/\sqrt{\Omega})$ into account. Since their exponents $W_2(x) - W_2(x_e)$ and $\Phi(x) - \Phi(x_e)$ agree locally up to second order, their difference is of the form

$$[W_2 - W_2(x_e)] - [\Phi - \Phi(x_e)] = [W_2'''(\xi) - \Phi'''(\xi)]\frac{(x - x_e)^3}{6}, \tag{60}$$

for some ξ. We thus deduce that $\Omega([W_2(x) - W_2(x_e)] - [\Phi(x) - \Phi(x_e)]) = \mathcal{O}(1/\sqrt{\Omega})$, from which it follows that

$$p_2 - p_\infty \sim \left[\frac{2}{\bar{\lambda}(x) + \bar{\mu}(x)} \mathrm{e}^{\Omega[W_2(x) - W_2(x_e) - \Phi(x) + \Phi(x_e)]} - \frac{1}{\sqrt{\bar{\lambda}(x)\bar{\mu}(x)}}\right] \times x_e \sqrt{\frac{\Omega\Lambda}{2\pi}} \mathrm{e}^{\Omega[\Phi(x) - \Phi(x_e)]} \tag{61a}$$

$$\sim \left[\frac{2}{\bar{\lambda}(x) + \bar{\mu}(x)} - \frac{1}{\sqrt{\bar{\lambda}(x)\bar{\mu}(x)}}\right] x_e \sqrt{\frac{\Omega\Lambda}{2\pi}} \mathrm{e}^{\Omega[\Phi(x) - \Phi(x_e)]} \tag{61b}$$

$$= -\left[\frac{\left(\sqrt{\bar{\lambda}(x)} - \sqrt{\bar{\mu}(x)}\right)^2}{\sqrt{\bar{\lambda}(x)\bar{\mu}(x)}\left[\bar{\lambda}(x) + \bar{\mu}(x)\right]}\right] x_e \sqrt{\frac{\Omega\Lambda}{2\pi}} \mathrm{e}^{\Omega[\Phi(x) - \Phi(x_e)]}. \tag{61c}$$

Fascinatingly, the difference between these two solutions essentially boils down to the difference between (the inverses of) the arithmetic and geometric means of $\bar{\lambda}$ and $\bar{\mu}$, encapsulated in the bracketed term in (61b). Since $\sqrt{\bar{\lambda}(x)} - \sqrt{\bar{\mu}(x)} \sim \Lambda\sqrt{x}(x - x_e)/2 = \mathcal{O}(1/\sqrt{\Omega})$, we conclude that

$$\|P_2 - P_\infty\|_n = \mathcal{O}\left(\Omega^{-1.5+1/(2n)}\right). \tag{62}$$

Finally, we compare P_∞ and the exact discrete solution $P_{X>0}$. The next order correction in WKB approximation is $O(1/\Omega)$, which suggests that

$$\|P_\infty - P_{X>0}\|_n = \mathcal{O}\left(\Omega^{-2+1/(2n)}\right). \tag{63}$$

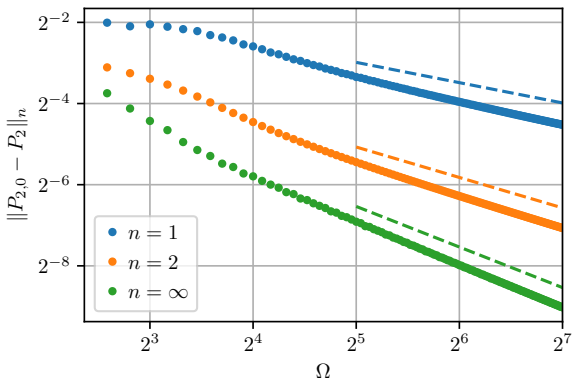
In fig. 5 we plot each of the continuum approximations compared with the exact discrete solution. We find good agreement with the error estimates derived above. Since each successive error estimate was smaller than the previous, we may telescope the errors and deduce the same error in comparison with the discrete solution $P_{X>0}$. For example, we see that

$$\|P_{2,0} - P_{X>0}\|_n \leq \|P_{2,0} - P_2\|_n + \|P_2 - P_\infty\|_n + \|P_\infty - P_{X>0}\|_n = \mathcal{O}\left(\Omega^{-1+1/(2n)}\right),$$

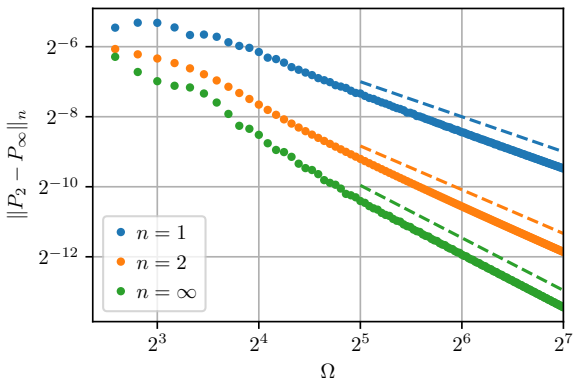
since the first term on the right-hand side of the inequality is dominant. Even though the asymptotic Fokker–Planck solution is the least accurate, it was invaluable in uncovering the structure that underpins the error estimates for all of the solutions.

To conclude, we find that the ESD contains a similar structure to the mean first-passage times: higher-order terms in the continuum master equation approximation are, formally, increasingly negligible, and yet persist in contributing to the leading-order exponent. However, in contrast to the first-passage time context, the normalisation renders these discrepancies vanishingly small in except in a boundary layer in the vicinity of the metastable equilibrium point. In this neighbourhood, a quantitative analysis reveals the asymptotic accuracy of . The Fokker–Planck approximation proves remarkably accurate at resolving the ESD of the discrete process, and provides valuable transient and structural information about the problem, while higher-order contributions make only an asymptotically small contribution.

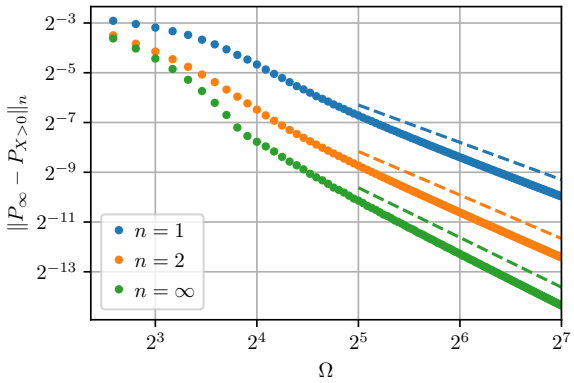
The analysis in this section relies only loosely on the particular forms of the birth and death rates (6). The matched asymptotic calculation relied on the assumption that $\bar{\lambda} - \bar{\mu}$ has a simple root, but this algebraic condition is precisely what provides the assumed metastable equilibrium, and so entails no loss of generality. Similarly, the non-integrability of the continuum ESDs arises from the aforementioned fact that, in the case of degenerate diffusion (that is, where $\bar{\lambda} + \bar{\mu} = 0$ at $x = 0$, as is the case for (6)) the no-flux and absorbing boundary conditions are formally equivalent. So while we were able to define the discrete ESD by changing an absorbing boundary into a reflecting boundary, in the continuum no such distinction can be made. Were we to consider a process with diffusion uniformly bounded away from zero, the non-integrability complication would not arise. As in § 2, we rely primarily on the WKB ansatz with $\mathcal{O}(1)$ coefficient and exponent. Therefore, we deduce that these calculations pertain more generally to one-dimensional birth–death processes, motivating further study in higher dimension and for a wider class of processes.



(a) Slope $-1 + 1/(2n)$ from (59).



(b) Slope $-1.5 + 1/(2n)$ from (62).



(c) Slope $-2 + 1/(2n)$ from (63).

Figure 5: Discrepancies between different QSDs. Dashed lines represent the predicted convergence. The comparison is done in the original scaling, hence the capital letters, by undoing (12).

4 Conclusions

Following observations of exponentially large errors in mean first-passage time problems for birth–death processes, it has been suggested that the continuum approximation is only valid under strict conditions on the birth and death rates, as well as the system size being large but not too large, resulting in uncertainty surrounding the accuracy and applicability of the technique.

In this paper we present a quantitative argument explaining the cause of the discrepancy. The Fokker–Planck approximation is arrived at via a second-order truncation of a series expansion of the master equation. In mean first-passage time problems, every term in the series makes an order unity contribution (with respect to the system size limit) to the first-passage time, therefore low-order truncation should be expected to lead to exponential error. However, the terms in the series diminish rapidly (at a factorial rate, with respect to the truncation order), and so corrections to the Fokker–Planck approximation within the continuum framework are available and tractable. We argue that this is the appropriate perspective from which to view the approximation error: Fokker–Planck is a coarse member in a hierarchy of convergent approximations that bridge the discrete and continuous processes. Therefore, we expect it to be an invaluable tool, even for first-passage time problems, if only qualitatively. In instances where more quantitative agreement is necessary, the appropriate higher-order truncations may be instated. Couching the problem in the continuum provides a direct way to bring to bear the extensive tools of differential calculus, alongside numerical techniques that may offer a more flexible discretisation than available via the original discrete process.

To emphasise how these higher-order terms contribute in other problems, we analyse the distributions

of the same birth–death process. Despite an analogous structure, the Fokker–Planck approximation is exceptionally accurate, even for moderate system sizes, and provides a quantification of where deterministic versus stochastic contributions dominate the dynamics. We thereby confirm that the exponential magnitude of previously observed inaccuracies pertain solely to problems of specific structure as revealed through the asymptotic analysis of the exponentially long first-passage times.

Our analysis advocates for a broad interpretation of "continuum approximations", to include approximations of the probability density even in the absence of an underlying stochastic process. We demonstrate how these provide indispensable insight, and deduce that, despite the prevalence of the second-order truncation, higher orders may be desirable in certain contexts, such as the first-passage time problems studied in this paper. With this in mind, we are prompted to reconsider sweeping suggestions that the continuum approach is "delicate" or only applicable in "very special situations".

Most of the analysis is performed for a fairly generic birth–death process, assuming only the existence of a metastable state. The specific birth and death rates analysed in [5] are employed primarily for the purposes of illustration. We thus conclude that our observations should generalise broadly to one-dimensional birth–death processes. This raises the question of how far we can extend these observations. Looking forward, it would be intriguing to explore whether or not the dimensionality or birth–death structure play a significant role. To what extent do these results hold in higher dimension? How do they depend on the structure of the dynamical system? Furthermore, we focus exclusively on the *mean* first-passage time, and it would be interesting to explore how these observations pertain to higher-order moments and other statistics, considering that distributions in real-world applications may not be well characterised by their mean [19].

The practitioner should be aware of the limits of any tool used, and continuum approximations of discrete Markov processes are no exception. While continuum approximations will incur some approximation error (see also [2, 12]), we hope that the insights revealed in this study regarding both the source and remedy of error, inspire further work in these, and other, directions.

A Laplace's method

In this appendix, we apply Laplace's method to obtain an asymptotic approximation of integrals of the form $\int g(\xi)\,\mathrm{e}^{\Omega f(\xi)}\,\mathrm{d}\xi$ in the limit as $\Omega \to \infty$, assuming that both $g(x)$ and $f(x)$ are of order unity. The central idea is that, since the exponent is large, the dominant contributions to the integral come from neighbourhoods of global maxima of f, as all other regions are exponentially smaller. We demonstrate two cases from the main text: first, where f has a global maximum on the boundary of the domain (with non-zero derivative), and second, where f has a global maximum in the domain interior.

A.1 Boundary maximum

We begin with (16) and (17) for the mean first-passage time with truncation order N, whereby

$$\tau_N = \int_0^x \frac{\mathrm{d}}{\mathrm{d}\xi}\tau_k(\xi)\,\mathrm{d}\xi \sim \int_0^x c_N(\xi)\,\mathrm{e}^{\Omega V_N(\xi)}\,\mathrm{d}\xi, \tag{64}$$

with $V_N' < 0$ on $x \in [0, x_e)$. Therefore, the dominant contribution is expected to be in a neighbourhood of the origin, and we split the domain of integration via

$$\tau_N \sim \int_0^\delta c_N(\xi)\,\mathrm{e}^{\Omega V_N(\xi)}\,\mathrm{d}\xi + \int_\delta^x c_N(\xi)\,\mathrm{e}^{\Omega V_N(x)}\,\mathrm{d}\xi, \tag{65}$$

for some $\delta \ll 1$ which we will determine. Expanding locally allows us to evaluate each integral contribution as follows. Near the origin, and assuming that

$$\Omega\delta \gg 1 \qquad \text{and} \qquad \Omega\delta^2 \ll 1, \tag{66}$$

we find that

$$\begin{aligned}\int_0^\delta c_N(\xi)\,\mathrm{e}^{\Omega V_N(\xi)}\,\mathrm{d}\xi &= \int_0^\delta [c_N(0)+\mathcal{O}(\delta)]\,\mathrm{e}^{\Omega[V_N(0)+\xi V_N'(0)+\mathcal{O}(\delta^2)]}\,\mathrm{d}\xi \\ &\sim c_N(0)\,\mathrm{e}^{\Omega V_N(0)}\int_0^\delta \mathrm{e}^{\Omega\xi V_N'(0)}\,\mathrm{d}\xi \\ &= -\frac{c_N(0)}{\Omega V_N'(0)}\,\mathrm{e}^{\Omega V_N(0)}\left[1-\mathrm{e}^{\Omega\delta V_N'(0)}\right] \\ &\sim -\frac{c_N(0)}{\Omega V_N'(0)}\,\mathrm{e}^{\Omega V_N(0)}.\end{aligned} \tag{67}$$

Away from the origin, since $V_N' < 0$, we find that the integral is exponentially smaller than the contribution (67), admitting the bound

$$\int_\delta^x c_N(\xi)\,\mathrm{e}^{\Omega V_N(x)}\,\mathrm{d}\xi = \mathcal{O}\left(\mathrm{e}^{\Omega V_N(\delta)}\right) = \mathcal{O}\left(\mathrm{e}^{\Omega[V_N(0)+\delta V_N'(0)+\mathcal{O}(\delta^2)]}\right). \tag{68}$$

It thus suffices to find δ satisfying conditions (66), for example, $\delta = \Omega^{-\alpha}$ for $\alpha \in (1/2, 1)$. Then, substituting (67) and (68) into (64), we deduce that

$$\tau_N \sim -\frac{c_N(0)}{\Omega V_N'(0)}\,\mathrm{e}^{\Omega V_N(0)}, \tag{69}$$

from which we deduce that

$$\log(\tau_N)/\Omega \sim V_N(0). \tag{70}$$

A.2 Interior maximum

In this subsection, we begin with (42), whereby we have the probability density p_2 normalised by the constant $\hat{c}_2$, such that

$$\frac{1}{\hat{c}_2} = \int_\epsilon^1 \frac{\mathrm{e}^{\Omega W_2(\xi)}}{\bar{\lambda}(\xi)+\bar{\mu}(\xi)}\,\mathrm{d}\xi\,\mathrm{d}\xi, \tag{71}$$

where $W_2(x)$ obtains a unique maximum at $x = x_e$. We split the domain of integration, isolating a neighbourhood of the maximum $[x_e-\delta, x_e+\delta]$ for $\delta \ll 1$ that we will describe. Assuming

$$\Omega\delta^2 \gg 1 \qquad \text{and} \qquad \Omega\delta^3 \ll 1, \tag{72}$$

we see that

$$\int_{x_e-\delta}^{x_e+\delta} \frac{\mathrm{e}^{\Omega W_2(\xi)}}{\bar{\lambda}(\xi)+\bar{\mu}(\xi)}\,\mathrm{d}\xi\,\mathrm{d}\xi = \int_{x_e-\delta}^{x_e+\delta} \frac{1+\mathcal{O}(\delta)}{\bar{\lambda}(x_e)+\bar{\mu}(x_e)}\,\mathrm{e}^{\Omega[W_2(x_e)+W_2''(x_e)(\xi-x_e)^2/2+\mathcal{O}(\delta^3)]}\,\mathrm{d}\xi \tag{73a}$$

$$\sim \frac{\mathrm{e}^{\Omega W_2(x_e)}}{\bar{\lambda}(x_e)+\bar{\mu}(x_e)}\int_{x_e-\delta}^{x_e+\delta}\mathrm{e}^{\Omega W_2''(x_e)(\xi-x_e)^2/2}\,\mathrm{d}\xi \tag{73b}$$

$$= \frac{\mathrm{e}^{\Omega W_2(x_e)}}{\bar{\lambda}(x_e)+\bar{\mu}(x_e)}\sqrt{\frac{2}{-\Omega W_2''(x_e)}}\int_{-\zeta_0}^{\zeta_0}\mathrm{e}^{-\zeta^2}\,\mathrm{d}\zeta \tag{73c}$$

$$\sim \frac{\mathrm{e}^{\Omega W_2(x_e)}}{\bar{\lambda}(x_e)+\bar{\mu}(x_e)}\sqrt{\frac{2\pi}{-\Omega W_2''(x_e)}}, \tag{73d}$$

where $\zeta_0 = \delta\sqrt{-\Omega W_2''(x_e)/2}$. The deduction of (73d) follows by extending the range of integration to $\zeta = \pm\infty$, which introduces exponentially small errors assuming conditions (72) are met.

Away from the maximum, $W_2'(x) \geq 0$ on $x < x_e$, allowing us to bound the integral via

$$\int_{\epsilon}^{x_e-\delta} \frac{\mathrm{e}^{\Omega W_2(\xi)}}{\bar{\lambda}(\xi)+\bar{\mu}(\xi)}\,\mathrm{d}\xi = \mathcal{O}\left(\mathrm{e}^{\Omega W_2(x_e-\delta)}\right) = \mathcal{O}\left(\mathrm{e}^{\Omega[W_2(x_e)-\delta^2 W_2''(x_e)/2+\mathcal{O}(\delta^3)]}\right), \tag{74}$$

which is exponentially smaller than the contributions near the maximum (73), when conditions (72) are satisfied. Similarly, $W_2'(x) \leq 0$ on $x > x_e$, and the associated integral contribution is exponentially smaller than contribution (73).

Finally, conditions (72) are satisfied by, for example, $\delta = \Omega^{-\alpha}$ for $\alpha \in (1/3, 1/2)$. We therefore conclude that

$$\hat{c}_2 \sim \left[\bar{\lambda}(x_e)+\bar{\mu}(x_e)\right]\mathrm{e}^{-\Omega W_2(x_e)}\sqrt{\frac{-\Omega W_2''(x_e)}{2\pi}}. \tag{75}$$

B First-order calculation for WKB coefficient

The steady form of equation (13) for $N = \infty$ is given by

$$0 = \sum_{k=1}^{\infty} \frac{1}{\Omega^{k-1}k!}\frac{\mathrm{d}^k}{\mathrm{d}x^k}\left[b_k(x)p_\infty(x)\right], \qquad b_k(x) = \begin{cases} \bar{\mu}(x)-\bar{\lambda}(x), & k \text{ odd}, \\ \bar{\mu}(x)+\bar{\lambda}(x), & k \text{ even}. \end{cases} \tag{76}$$

Taking the WKB ansatz (and suppressing the subscript $N = \infty$ to simplify the notation)

$$p_\infty \sim (c(x)+o(1))\,\mathrm{e}^{\Omega W(x)}, \tag{77}$$

we seek to express the summand term $\mathrm{d}^k(b_k p)/\mathrm{d}x^k$ in successive powers of Ω. We employ two main results. First, Liebniz's product rule, whereby

$$\frac{\mathrm{d}^k}{\mathrm{d}x^k}[f(x)g(x)] = \sum_{i=0}^{k}\binom{k}{i}f^{(k-i)}(x)g^{(i)}(x). \tag{78}$$

Second, Faà di Bruno's formula [18], namely

$$\frac{\mathrm{d}^k}{\mathrm{d}x^k}f(g(x)) = \sum_{m_1+2m_2+\cdots+km_k=k}\frac{k!}{m_1!\cdots m_k!}f^{\left(\sum_{i=1}^k m_i\right)}(g(x))\prod_{j=1}^{k}\left(\frac{g^{(j)}(x)}{j!}\right)^{m_j}, \tag{79}$$

where the outer sum is taken over all k-tuples of natural numbers satisfying the constraint $m_1 + 2m_2 + \cdots + km_k = k$. Using $f(x) = \mathrm{e}^{\Omega x}$ and $g(x) = W(x)$, we see that in (79) the order of magnitude of each summand is determined by the order of the derivative of f. The lowest order contribution comes from the k-tuple $(k, 0, \ldots, 0)$, the first order contribution comes from $(k-2, 1, 0\ldots, 0)$, while contributions from all other admissible k-tuples are of higher order. It then follows from (79) that

$$\frac{\mathrm{d}^k}{\mathrm{d}x^k}\mathrm{e}^{\Omega W(x)} = \mathrm{e}^{\Omega W}\left[\Omega^k(W')^k + \Omega^{k-1}\frac{k(k-1)}{2}(W')^{k-2} + \mathcal{O}\left(\Omega^{k-2}\right)\right]. \tag{80}$$

Then from Leibniz's formula (78) and (80) we see that

$$\begin{aligned}\frac{\mathrm{d}^k}{\mathrm{d}x^k}p_\infty &= \sum_{i=0}^{k}\binom{k}{i}c^{(k-i)}\frac{\mathrm{d}^i}{\mathrm{d}x^i}\mathrm{e}^{\Omega W(x)} \\ &= \mathrm{e}^{\Omega W(x)}\left\{\Omega^k\left[(W')^k c\right]\right. \\ &\qquad \left. + \Omega^{k-1}\left[\frac{k(k-1)}{2}(W')^{k-2}W''c + k(W')^{k-1}c'\right] + \mathcal{O}\left(\Omega^{k-2}\right)\right\}.\end{aligned}$$

An additional application of Leibniz's formula produces

$$\begin{aligned}\frac{\mathrm{d}^k}{\mathrm{d}x^k}[b_k p_\infty] &= \sum_{i=0}^{k} \binom{k}{i} b_k^{(k-i)} \frac{\mathrm{d}^i}{\mathrm{d}x^i} p_\infty \\ &= \mathrm{e}^{\Omega W(x)} \left\{ \Omega^k \left[(W')^k b_k c \right] \right. \\ &\qquad \left. + \Omega^{k-1} \left[\frac{k(k-1)}{2} (W')^{k-2} W'' b_k c + k (W')^{k-1} (b_k c)' \right] + \mathcal{O}\left(\Omega^{k-2}\right) \right\}. \end{aligned} \tag{81}$$

Substituting (81) into (76), at leading order $\mathcal{O}(\Omega)$ we recover equation (56) governing W, as discussed in the main text. At the next order $\mathcal{O}(1)$, we obtain the equation governing c, namely

$$0 = \sum_{k=2}^{\infty} \frac{(W')^{k-2}}{(k-2)!} \frac{W'' b_k c}{2} + \sum_{k=1}^{\infty} \frac{(W')^{k-1}}{(k-1)!} (b_k c)' \tag{82a}$$

$$= \frac{W'' b_2 c}{2} \sum_{k=0}^{\infty} \frac{(W')^{2k}}{(2k)!} + \frac{W'' b_1 c}{2} \sum_{k=1}^{\infty} \frac{(W')^{2k-1}}{(2k-1)!} + (b_1 c)' \sum_{k=0}^{\infty} \frac{(W')^{2k}}{(2k)!} + (b_2 c)' \sum_{k=1}^{\infty} \frac{(W')^{2k-1}}{(2k-1)!} \tag{82b}$$

$$= \left[\frac{W'' b_2 c}{2} + (b_1 c)' \right] \cosh(W') + \left[\frac{W'' b_1 c}{2} + (b_2 c)' \right] \sinh(W'). \tag{82c}$$

Upon rearranging equation (82c) and integrating once, we obtain

$$c(x) = \frac{\hat{c}_\infty}{\sqrt{\bar{\lambda}(x)\bar{\mu}(x)}}, \tag{83}$$

for a normalising constant $\hat{c}_\infty$. Combining the results, the leading-order WKB approximation is given by

$$p_\infty(x) \sim \frac{\hat{c}_\infty \, \mathrm{e}^{\Omega \Phi(x)}}{\sqrt{\bar{\lambda}(x)\bar{\mu}(x)}}, \qquad \hat{c}_\infty \sim (1 - 1/\Lambda) \, \mathrm{e}^{-\Omega \Phi(x_e)} \sqrt{\frac{\Omega \Lambda}{2\pi}}, \tag{84}$$

where the constant $\hat{c}_\infty$ admits an asymptotic approximation via Laplace's method (see appendix A). In principle, one could proceed to higher orders to determine asymptotic corrections to c in inverse powers of Ω.

Acknowledgements

I would like to thank Prof. J. Frédéric Bonnans and Dr. Jakob Ruess for their insightful advice and helpful comments. This work was supported by the Inria Project Lab CoSy.

References

[1] M. Assaf, B. Meerson, and P. V. Sasorov, *Large fluctuations in stochastic population dynamics: momentum-space calculations*, Journal of Statistical Mechanics: Theory and Experiment, 2010 (2010), p. P07018, https://doi.org/10.1088/1742-5468/2010/07/p07018.

[2] F. Baras, M. M. Mansour, and J. E. Pearson, *Microscopic simulation of chemical bistability in homogeneous systems*, The Journal of Chemical Physics, 105 (1996), pp. 8257–8261, https://doi.org/10.1063/1.472679.

[3] G. E. P. Box, *Science and statistics*, Journal of the American Statistical Association, 71 (1976), pp. 791–799, https://doi.org/10.1080/01621459.1976.10480949.

[4] P. C. Bressloff, *Stochastic Processes in Cell Biology*, Springer International Publishing, 2014, https://doi.org/10.1007/978-3-319-08488-6.

[5] C. R. Doering, K. V. Sargsyan, and L. M. Sander, *Extinction times for birth-death processes: Exact results, continuum asymptotics, and the failure of the Fokker–Planck approximation*, Multiscale Modeling & Simulation, 3 (2005), pp. 283–299.

[6] S. Engblom, *Computing the moments of high dimensional solutions of the master equation*, Applied Mathematics and Computation, 180 (2006), pp. 498 – 515, https://doi.org/10.1016/j.amc.2005.12.032.

[7] A. Erdélyi, *Asymptotic expansions*, Dover Publications, 1956.

[8] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, Springer Series in Synergetics, Springer Berlin Heidelberg, third ed., 2009.

[9] C. W. Gardiner and S. Chaturvedi, *The poisson representation. I. A new technique for chemical master equations*, Journal of Statistical Physics, 17 (1977), pp. 429–468, https://doi.org/10.1007/BF01014349.

[10] M. A. Gibson and J. Bruck, *Efficient exact stochastic simulation of chemical systems with many species and many channels*, The Journal of Physical Chemistry A, 104 (2000), pp. 1876–1889, https://doi.org/10.1021/jp993732q.

[11] C. S. Gillespie, *Moment-closure approximations for mass-action models*, IET Systems Biology, 3 (2009), pp. 52–58, https://doi.org/10.1049/iet-syb:20070031.

[12] D. T. Gillespie, *The chemical Langevin equation*, The Journal of Chemical Physics, 113 (2000), pp. 297–306, https://doi.org/10.1063/1.481811.

[13] J. Grasman, *The expected extinction time of a population within a system of interacting biological populations*, Bulletin of Mathematical Biology, 58 (1996), pp. 555 – 568, https://doi.org/https://doi.org/10.1016/0092-8240(95)00361-4.

[14] R. Grima, P. Thomas, and A. V. Straube, *How accurate are the nonlinear chemical Fokker-Planck and chemical Langevin equations?*, J. Chem. Phys., 135 (2011), p. 084103, https://doi.org/10.1063/1.3625958.

[15] E. J. Hinch, *Perturbation Methods*, Cambridge Texts in Applied Mathematics, Cambridge University Press, 1991, https://doi.org/10.1017/CBO9781139172189.

[16] N. G. v. Kampen, *A power series expansion of the master equation*, Can. J. Phys., 39 (1961), pp. 551–567, https://doi.org/10.1139/p61-056.

[17] J. K. Kim, K. Josić, and M. R. Bennett, *The validity of quasi-steady-state approximations in discrete stochastic simulations*, Biophys. J., 107 (2014), pp. 783–793, https://doi.org/10.1016/j.bpj.2014.06.012.

[18] S. G. Krantz and H. R. Parks, *A primer of real analytic functions*, Birkhäuser, 2 ed., 2002, https://doi.org/10.1007/978-0-8176-8134-0.

[19] D. Lunz, G. Batt, J. Ruess, and J. F. Bonnans, *Beyond the chemical master equation: stochastic chemical kinetics coupled with auxiliary processes*, (2020). https://hal.inria.fr/hal-02991103.

[20] B. J. Matkowsky, Z. Schuss, C. Knessl, C. Tier, and M. Mangel, *First passage times for processes governed by master equations*, in Fluctuations and Sensitivity in Nonequilibrium Systems, W. Horsthemke and D. K. Kondepudi, eds., Berlin, Heidelberg, 1984, Springer Berlin Heidelberg, pp. 19–36.

[21] S. Méléard and D. Villemonais, *Quasi-stationary distributions and population processes*, Probab. Surveys, 9 (2012), pp. 340–410, https://doi.org/10.1214/11-PS191.

[22] V. Méndez, M. Assaf, A. Masó-Puigdellosas, D. Campos, and W. Horsthemke, *Demographic stochasticity and extinction in populations with Allee effect*, Phys. Rev. E, 99 (2019), p. 022101, https://doi.org/10.1103/PhysRevE.99.022101.

[23] R. Milo, *What is the total number of protein molecules per cell volume? A call to rethink some published values*, BioEssays, 35 (2013), pp. 1050–1055, https://doi.org/10.1002/bies.201300066.

[24] L. Pauling, *General Chemistry*, Dover Books on Chemistry, Dover Publications, 1988.

[25] R. F. Pawula, *Approximation of the linear Boltzmann equation by the Fokker–Planck equation*, Phys. Rev., 162 (1967), pp. 186–188, https://doi.org/10.1103/PhysRev.162.186.

[26] H. Risken, *Fokker–Planck Equation*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1996, https://doi.org/10.1007/978-3-642-61544-3_4.

[27] P. Thomas, H. Matuschek, and R. Grima, *How reliable is the linear noise approximation of gene regulatory networks?*, BMC Genomics, 14 (2013), p. S5, https://doi.org/10.1186/1471-2164-14-S4-S5.

[28] N. G. van Kampen, *Stochastic processes in physics and chemistry*, North Holland, third ed., 2007, https://doi.org/10.1016/B978-0-444-52965-7.X5000-4.

[29] X. Yu and X.-Y. Li, *Applications of wkb and Fokker–Planck methods in analyzing population extinction driven by weak demographic fluctuations*, Bulletin of Mathematical Biology, 81 (2019), pp. 4840–4855, https://doi.org/10.1007/s11538-018-0483-6.